WHAT IS CLAIMED IS:

1. A computer-implemented method of generating concept units from user search queries, the method comprising:

receiving a plurality of queries, each query comprising a string of one or more words;

tokenizing each query string to produce one or more tokens for each query, wherein said tokens for said queries form an initial set of units;

combining units from the initial set of units that appear adjacent each other in a query to form a second set of units;

validating the second set of units;

repeating the steps of combining and validating one or more times using the second set of units in place of the initial set of units until a convergence condition is satisfied, wherein a final set of units is formed once the convergence condition has been satisfied; and

storing the final set of units to a memory.

2. The method of claim 1, wherein receiving includes receiving one or more query log files, each query log file including a plurality of queries.

3. The method of claim 2, further comprising consolidating the plurality of queries from the one or more query log files into a single consolidated query file.

4. The method of claim 3, wherein consolidating includes removing duplicates of queries and incrementing a count associated with each individual query each time a duplicate of said individual query is removed, wherein the consolidated file includes a list of individual queries and counts associated therewith.

5. The method of claim 3, wherein the received query log files include query log files for each day of a week, and wherein consolidating includes forming a single consolidated query file including queries for the week.

6. The method of claim 1, further including generating unit extensions using the final set of units

7. The method of claim 6, wherein generating unit extensions includes identifying units that are subsets of other units.

8. The method of claim 6, further including storing the unit extensions to the memory.

9. The method of claim 1, further including generating unit associations using the final set of units.

10. The method of claim 9, wherein generating unit associations includes identifying units that are associated with other units.

11. The method of claim 10, further including storing the unit associations to the memory.

12. The method of claim 10, wherein identifying associated units includes determining which units appear in queries with other units.

13. The method of claim 1, further comprising generating unit alternatives after the convergence condition has been satisfied.

14. The method of claim 13, wherein generating unit alternatives includes determining whether an edit distance between two units in the final set of units is smaller than a threshold value, and if so, comparing the relative frequencies of the two units.

15. The method of claim 1, further comprising:

generating unit extensions using the final set of units;

generating unit associations using the final set of units; and

generating unit alternatives using the final set of units.

16. The method of claim 15, further including storing the unit extensions, the unit associations and the unit alternatives to the memory.

17. The method of claim 15, wherein generating unit extensions includes identifying units that are subsets of other units, wherein generating unit associations includes identifying units that are associated with other units, and wherein generating unit alternatives includes determining whether an edit distance between two units in the final set of units is smaller than a threshold value, and if so, comparing the relative frequencies of the two units.

18. The method of claim 1, wherein validating includes for each combined unit in the second set of units, comparing a frequency of occurrence of the combined unit with a frequency of occurrence of each constituent unit in the combined unit.

19. The method of claim 1, wherein the convergence condition includes a threshold value, wherein the convergence condition is satisfied if a change in the number of units in the two second set of units between successive steps of combining and validating is smaller than or equal to the threshold value.

20. The method of claim 1, further including:

receiving an individual query from a user;

identifying one or more units in the individual query; and

determining one or more suggestions to provide to the user responsive to the query using one or more of the unit extensions, unit associations and unit alternatives stored in the memory in association with the one or more units identified in the individual query.

21. A system for generating concept units from user search queries, the system comprising:

a memory unit; and

a processing module configured to receive one or more query log files, each query log file including a plurality of queries, each query including a string of one or more words, and wherein the processing module is further configured to:

tokenize each query from the query log files to produce an initial set of units; and thereafter, iteratively, until a convergence condition is satisfied:

combine units from the initial set of units that appear adjacent each other in a query to form a second set of units; and

validate the second set of units, wherein the second set of units is used for each iteration; and

once the convergence condition has been satisfied, store a final set of units to the memory unit.

22. The system of claim 21,further including one or more query log file sources for providing the query log files.

23. The system of claim 21, wherein the processing module is further configured to:

generate unit extensions using the final set of units;

generate unit associations using the final set of units;

generate unit alternatives using the final set of units; and

store the unit extensions, unit associations and unit alternatives to the memory unit in association with the final set of units.

24. The system of claim 21, wherein the received query log files include query log files for each day of a week, and wherein the processing module is further configured to consolidate the query log files into a single consolidated query file consisting of queries for the week.

25. The system of claim 24, wherein the processing module consolidates by removing duplicates of queries and incrementing a count associated with each individual query each time a duplicate of said individual query is removed, wherein the consolidated file includes a list of individual queries and counts associated therewith.

26. The system of claim 25, wherein the processing module determines a frequency of occurrence for each unit using the counts associated with the queries, and wherein the processing modules stores the unit frequencies to the memory unit in association with the final set of units.

27. The system of claim 21, wherein the memory unit and processing module are implemented in a search server device in a network.

28. A computer readable medium including code for causing a processor to generate concept units from a plurality of user search queries, each query comprising a string of one or more words ,wherein the code includes instructions to:

a) tokenize each query string to produce one or more tokens for each query, wherein said tokens for said queries form an initial set of units;

b) combine units from the initial set of units that appear adjacent each other in a query to form a second set of units;

c) validate the second set of units;

d) repeat b) and c) one or more times using the second set of units in place of the initial set of units until a convergence condition is satisfied, wherein a final set of units is formed once the convergence condition has been satisfied; and

store the final set of units to a memory module.

29. The computer-readable medium of claim 28, wherein the code further includes instructions to:

generate unit extensions using the final set of units;

generate unit associations using the final set of units;

generate unit alternatives using the final set of units; and

store the unit extensions, unit associations and unit alternatives to the memory module in association with the final set of units.

30. The computer-readable medium of claim 29, wherein the instructions to generate unit extensions includes instructions to identify units that are subsets of other units, wherein the instructions to generate unit associations includes instructions to identify units that are associated with other units, and wherein the instructions to generate unit alternatives includes instructions to determine whether an edit distance between two units in the final set of units is smaller than a threshold value, and if so, compare the relative frequencies of the two units.

31. The method of claim 1, wherein each word comprises one or a plurality of alphanumeric characters.